WHAT IS CLAIMED IS:

1. A method of segmenting words into component parts, the method comprising:

determining mutual information scores for graphoneme units, each graphoneme unit comprising at least on letter in the spelling of a word;

using the mutual information scores to combine graphoneme units into a larger graphoneme unit; and

segmenting words into component parts to form a sequence of graphonemes.

2. The method of claim 1 wherein combining graphonemes comprises combining the letters of each graphoneme to produce a sequence of letters for the larger graphoneme unit and combining the phones of each graphoneme to produce a sequence of phones for the larger graphoneme unit.

3. The method of claim 1 further comprising using the segmented words to generate a model.

4. The method of claim 3 wherein the model describes the probability of a graphoneme unit given a context within a word.

5. The method of claim 4 further comprising using the model to determine a pronunciation of a word given the spelling of the word.

6. The method of claim 1 wherein using the mutual information scores comprises summing at least two mutual information scores determined for a single larger graphoneme unit to form a strength.

7. A computer-readable medium having computer-executable instructions for performing steps comprising:

determining mutual information scores for pairs of graphoneme units found in a set of words, each graphoneme unit comprising at least one letter;

combining the graphoneme units of one pair of graphonome units to form a new graphoneme unit based on the mutual information scores; and

identifying a set of graphoneme units for a word based in part on the new graphoneme unit.

8. The computer-readable medium of claim 7 wherein combining the graphoneme units comprises combining the letters of the graphoneme units to form a sequence of letters for the new graphoneme unit.

9. The computer-readable medium of claim 8 wherein combining the graphoneme units further comprises combining the phones of the graphoneme

units to form a sequence of phones for the new gaphoneme unit.

10. The computer-readable medium of claim 7 further comprising identifying a set of graphonemes for each word in a dictionary.

11. The computer-readable medium of claim 10 further comprising using the sets of graphonemes identified for the words in the dictionary to train a model.

12. The computer-readable medium of claim 11 wherein the model describes the probability of a graphoneme unit appearing in a word.

13. The computer-readable medium of claim 12 wherein the probability is based on at least one other graphoneme unit in the word.

14. The computer-readable medium of claim 11 further comprising using the model to determine a pronunciation for a word given the spelling of the word.

15. The computer-readable medium of claim 7 wherein combining graphoneme units based on the mutual information score comprises summing at least two mutual information scores associated with a new graphoneme unit.

16. A method of segmenting a word into syllables, the method comprising:

segmenting a set of words into phonetic syllables using mutual information scores;

using the segmented set of words to train a syllable n-gram model; and

using the syllable n-gram model to segment a phonetic representation of a word into syllables via forced alignment.

17. A method of segmenting a word into morphemes, the method comprising:

segmenting a set of words into morphemes using mutual information scores;

using the segmented set of words to train a morpheme n-gram model; and

using the morpheme n-gram model to segment a word into morphemes via forced alignment.